

WHAT IS CLAIMED IS:

- 1 1. A method for processing anchor text, comprising:
2 forming a set of anchors that point to a target document;
3 grouping together anchors with same anchor text;
4 computing information for each group; and
5 generating context information for the target document based on the computed
6 information.
- 1 2. The method of claim 1, further comprising:
2 determining a language of each document in a collection of documents;
3 determining rank of each document in the collection of documents; and
4 determining a proximity class of each document in the collection of documents.
- 1 3. The method of claim 1, further comprising:
2 determining a predominant language in the set of anchors; and
3 pruning anchors from the set that are not in the predominant language.
- 1 4. The method of claim 1, further comprising:
2 pruning anchors from the set that include at least a portion of a path to the target
3 document.
- 1 5. The method of claim 1, further comprising:
2 pruning anchors based on a configurable set of words.
- 1 6. The method of claim 1, wherein computing information further comprises:
2 computing a weighted sum of occurrences for anchor text for anchors in each
3 group.
- 1 7. The method of claim 1, wherein computing information further comprises:

2 computing an accumulated rank for each group.

1 8. The method of claim 1, wherein computing information further comprises:
2 computing a linguistic score for each group.

1 9. Them method of claim 1, wherein computing information further
2 comprises:
3 generating a relevance score for each group.

1 10. A computer system including logic for processing anchor text,
2 comprising:
3 forming a set of anchors that point to a target document;
4 grouping together anchors with same anchor text;
5 computing information for each group; and
6 generating context information for the target document based on the computed
7 information.

1 11. The computer system of claim 10, wherein the logic further comprises:
2 determining a language of each document in a collection of documents;
3 determining rank of each document in the collection of documents; and
4 determining a proximity class of each document in the collection of documents.

1 12. The computer system of claim 10, wherein the logic further comprises:
2 determining a predominant language in the set of anchors; and
3 pruning anchors from the set that are not in the predominant language.

1 13. The computer system of claim 10, wherein the logic further comprises:
2 pruning anchors from the set that include at least a portion of a path to the target
3 document.

1 14. The computer system of claim 10, wherein the logic further comprises:
2 pruning anchors based on a configurable set of words.

1 15. The computer system of claim 10, wherein the logic for computing
2 information further comprises:
3 computing a weighted sum of occurrences for anchor text for anchors in each
4 group.

1 16. The computer system of claim 10, wherein the logic for computing
2 information further comprises:
3 computing an accumulated rank for each group.

1 17. The computer system of claim 10, wherein the logic for computing
2 information further comprises:
3 computing a linguistic score for each group.

1 18. Them computer system of claim 10, wherein the logic for computing
2 information further comprises:
3 generating a relevance score for each group.

1 19. An article of manufacture including a program for processing anchor text
2 in documents, wherein the program causes operations to be performed, the operations
3 comprising:
4 forming a set of anchors that point to a target document;
5 grouping together anchors with same anchor text;
6 computing information for each group; and
7 generating context information for the target document based on the computed
8 information.

1 20. The article of manufacture of claim 19, wherein the operations further
2 comprise:
3 determining a language of each document in a collection of documents;
4 determining rank of each document in the collection of documents; and
5 determining a proximity class of each document in the collection of documents.

1 21. The article of manufacture of claim 19, wherein the operations further
2 comprise:
3 determining a predominant language in the set of anchors; and
4 pruning anchors from the set that are not in the predominant language.

1 22. The article of manufacture of claim 19, wherein the operations further
2 comprise:
3 pruning anchors from the set that include at least a portion of a path to the target
4 document.

1 23. The article of manufacture of claim 19, wherein the operations further
2 comprise:
3 pruning anchors based on a configurable set of words.

1 24. The article of manufacture of claim 19, wherein the operations for
2 computing information further comprise:
3 computing a weighted sum of occurrences for anchor text for anchors in each
4 group.

1 25. The article of manufacture of claim 19, wherein the operations for
2 computing information further comprise:
3 computing an accumulated rank for each group.

1 26. The article of manufacture of claim 19, wherein the operations for
2 computing information further comprise:
3 computing a linguistic score for each group.

1 27. Them article of manufacture of claim 19, wherein the operations for
2 computing information further comprise:
3 generating a relevance score for each group.